

Debate sobre métodos frecuentistas vs bayesianos

L. C. Silva Ayçaguer¹ / A. Muñoz Villegas²

¹Vicerrectoría de Investigación y Posgrado. Instituto Superior de Ciencias Médicas. La Habana. Cuba.

²Department of Epidemiology. School of Hygiene and Public Health. Johns Hopkins University. Baltimore MD. USA.

³Dirección Xeral de Saúde Pública. Consellería de Sanidade e Servicios Sociais. Xunta de Galicia. España.

Moderador: E. Vázquez Fernández³

XVII Reunión Científica SEE. Santiago de Compostela 27-29 de octubre de 1999

Recibido: 26 de septiembre de 2000

Aceptado: 3 de octubre de 2000

(Debate on frequentist vs. Bayesian methods)

Moderador

Durante la primera reunión del Comité Científico, en la última reunión de la Sociedad Española de Epidemiología en Sevilla, me tocó trasladar al Comité una propuesta innovadora de José Luis Aboal: organizar un debate sobre los Métodos Frecuentistas y Bayesianos, dado el auge que últimamente está tomando este tema.

La forma de plantear el debate no deja de ser un poco teatral y busca que el ambiente sea relajado y que participe quien lo desee. No pretende ser una conferencia magistral en el sentido de marcar una clara diferencia entre los que estáis ahí sentados y los dos ponentes, sino que trata de ser abierta y con cierta dosis de improvisación. Decía que es teatral porque en realidad los dos provienen del campo frecuentista: Luis Carlos que va a «hacer de bayesiano», no lo es propiamente, si bien últimamente simpatiza claramente con esta corriente. El bayesianismo está entrando en nuestro campo con fuerza en la última década, aunque ha sido un viejo motivo de controversia entre los estadísticos durante todo este siglo, especialmente desde los años 30. Para ilustrar el auge del bayesianismo en la medicina, buscamos las citas en Medline a partir de 1966, tanto en el título como en el resumen de los artículos. En la primera década revisada, *bayesian* se citó en 29 títulos y ningún resumen, mientras que entre enero de 1995 y octubre de 1999 las citas en títulos llegaron a 235 y en los resúmenes a 577. Obviamente, alguien me podría decir, pues hay mucho epidemiólogo por aquí, que es necesario ajustar por el número total de citas en cada período. Éste no lo pude obtener, pero creo que los datos crudos ilustran lo que tratamos de señalar: un incremento del número de citaciones. Artículos como el de Freedman¹ y Lidford², en 1996 en *British Medical Journal*, o el de Greenland³ en *Epidemiology* del año 1998, o los más recientes de Goodman⁴ en *Annals of*

Internal Medicine de junio de 1999, sin olvidarnos de los de Bacallao^{5,6} en el año 1996 en *Medicina Clínica*, que también tenían una estructura similar a la que planteamos ahora, platoniana, de debate, nos ponen de manifiesto que es un tema que claramente está adquiriendo interés en nuestro campo de trabajo.

Tenemos otro dato más cercano, de esta mañana. La mesa dedicada a metodología, como bien sabéis, este año la han copado los métodos bayesianos de representación geográfica y no había un alfiler en la sala. Esto pone de manifiesto que en las Comunidades Autónomas este tema concreto de representación gráfica utilizando modelos bayesianos está entrando con fuerza.

Creo que muchos profesionales empezamos a preguntarnos si lo que hicimos en estos años estaba todo lo bien que pensábamos, o si estamos ante una moda y en el arranque del péndulo que no sabemos a dónde va. Personalmente son cuestiones que me inquietan y preocupan, por lo que espero este debate con expectación, ya que estoy seguro de que saldré con una opinión más clara que la que tengo ahora.

En cuanto a los ponentes, el tópico habitual de que no necesitan presentación, creo que en este caso es más que justificado. Ambos están a punto de obtener la tercera nacionalidad, el tercer pasaporte, de tanto venir a España, por suerte para todos nosotros. Pero de todas formas, por si hay alguien nuevo en este negocio que no los conozca, comento que Luis Carlos Silva Ayçaguer es Investigador Titular del Instituto Superior de Ciencias Médicas de La Habana. Álvaro Muñoz Villegas, es profesor del Departamento de Epidemiología de la Escuela de Salud Pública de la Universidad Johns Hopkins, de Baltimore.

Ayer por la noche estuvimos cenando y surgió una conversación, ya de pie, cuando nos retirábamos, que me resultó bastante inquietante, y me dije: «esto lo tengo que contar mañana». Ambos, no lo sabía a pesar

de que los conozco bastante bien, nacieron el mismo año. No voy a decir qué año, sólo diré que fue en la segunda mitad del siglo. Los dos se graduaron en matemáticas también el mismo año. Álvaro, como tiene un componente macondiano, asegura que fue el mismo día; Luis Carlos lo pone más en duda. Ninguno de los dos se dedica a las matemáticas, se dedican a la epidemiología; los dos tienen un hijo que se llama Daniel, las esposas de ambos se dedican a la bioestadística, ambos nacieron en Sudamérica, pero no residen en sus países de origen; a los dos les gusta España y en especial, no sé si será por cumplido, Galicia.... y no quise preguntar más porque no creo en las meigas.

Primero hablará Luis Carlos y responderá Álvaro; luego habrá una segunda intervención de Luis Carlos y otra de Álvaro. Por último, os quiero animar a que participéis en el debate que se abrirá al final de las presentaciones. Gracias.

Luis Carlos Silva

En primer lugar quiero aprovechar para agradecer muy sinceramente a los colegas de la Xunta de Galicia y a los Comités Organizador y Científico del Congreso la invitación que nos tiene aquí.

Voy a comenzar esta presentación haciendo una rápida valoración crítica en concordancia con una vocación que, en cierto sentido, me es consubstancial. Algunas de las ideas que expresaré se conectan con las que expuse oportunamente en mi libro «Cultura estadística e investigación científica en el campo de la salud: una mirada crítica»⁷.

En primer lugar, quisiera revisar rápidamente algunos aspectos que probablemente muchos de ustedes ya conozcan perfectamente, pero que de todas maneras vale la pena recordar.

El señor Ronald Fisher, en la década de los 20 propuso por primera vez valorar una hipótesis (H_0) a través de una observación concreta (d_0) así como la construcción de este número, el famoso « p » (p -value):

$$p = \text{Prob}(d \geq d_0 * H_0)$$

que es la probabilidad de observar algo mayor o igual que lo que objetivamente se observó, supuesto que sea válida la hipótesis que se valora. Él lo propuso como una medida de la discrepancia de los datos con la hipótesis. Ni más ni menos. Convocó a emplear esta medida, llamada a tener —junto con el resto de la información— un papel dentro del flujo orientando a la obtención de inferencias, aunque lo cierto es que nunca especificó formalmente cuál sería ese papel. No tardaron en producirse críticas a este enfoque. Me interesa subrayar una especialmente importante: se puede obtener

la misma p , tanto ante una diferencia observada muy grande, si el tamaño de muestra es pequeño, como ante una diferencia pequeña si dicho tamaño es muy grande. Esta ha sido una crítica recurrente, que se planteó con bastante fuerza entonces y que mantiene toda su vigencia hoy.

Recuérdese la ya muy conocida expresión de uno de los estadísticos más connotados de este siglo, el señor Savage, referida a las pruebas de hipótesis que se resume en que «cuando se sabe de antemano que las hipótesis de nulidad son falsas, el rechazo o la aceptación, simplemente es reflejo del tamaño de la muestra, y las pruebas no hacen por tanto contribución alguna a la ciencia»⁸. Esta realidad es válida para cualquier enfoque que trabaje con la p propuesta por Fisher y empleada luego por Neyman y Pearson para desarrollar su teoría.

¿Y cuán frecuentemente ocurre que se sabe de antemano que la hipótesis nula es falsa? Bakan decía que «casi nunca hay buenas razones para esperar que la hipótesis nula sea verdadera. ¿Por qué razón la media de los resultados de cierta prueba habría de ser exactamente igual al este que al oeste del río Mississippi? ¿Por qué deberíamos esperar que un coeficiente de correlación poblacional sea igual a 0.00? ¿Por qué esperar que la razón mujeres/hombres sea exactamente 50:50 en una comunidad dada?, o ¿por qué dos drogas habrán de producir exactamente el mismo efecto?»⁹. Lo cierto es que bastaría observar cualquier censo, para corroborar que la nulidad virtualmente nunca se produce en la naturaleza o en la sociedad.

El problema es bastante grave, porque significa que estamos planteando que vamos a tomar una decisión en función de cuán pequeño sea el valor de esa p , pero podemos conseguir que éste sea tan pequeño como queramos al tomar una decisión exógena a la realidad que estamos examinando (el tamaño de la muestra); entonces estamos ante un instrumento virtualmente condenado a ser intrascendente.

Esta realidad es tan chocante que incluso algunos defensores naturales de las pruebas de hipótesis, como lo hemos sido todos en una medida o en otra, han hecho planteamientos como el siguiente, que pertenece a Sacket; (vamos a cumplir el 20 aniversario de esta declaración, y lo subrayo porque da una idea de que estas preocupaciones no rigen desde los últimos tres años, ni nada por el estilo): «las muestras demasiado pequeñas pueden servir para no probar nada; las muestras demasiado grandes pueden servir para no probar nada»¹⁰.

Esto es muy interesante: no lo estaba diciendo desde una perspectiva crítica del enfoque convencional de las pruebas de significación; al decirlo, Sacket hace un reconocimiento objetivo de una notable endeblez que las afecta; pero es muy chocante porque lo que está sugiriendo es: «no miren mal la realidad, no miren una

parte pequeña de la realidad; de hacerlo, no van a poder sacar ninguna conclusión; pero tampoco la miren demasiado bien, porque si lo hacen, tampoco podrán sacar con esta herramienta conclusión alguna».

Lo peor que tiene esta afirmación de Sacket es que no es errónea. Si el tamaño de muestra es muy grande, se sabe de antemano lo que este instrumento (la prueba de hipótesis frecuentista) va a producir. Resulta obvio que una teoría con ese rasgo tiene una grave deficiencia: uno quisiera un procedimiento inferencial que, coherentemente con el sentido común, saque más partido a una muestra cuanto mayor ella sea.

Siguiendo con esta historia estadística, en los años 30 y como una reacción ante el planteamiento de Fisher, apareció por primera vez el tema de la hipótesis alternativa, ausente en el planteamiento original. Fue un planteamiento nuevo: se fijan unas tasas de error de falsos positivos y falsos negativos, como todos conocemos:

- α : es la probabilidad de rechazar la hipótesis nula cuando en realidad es cierta
- β : es la probabilidad de aceptarla cuando ella es falsa

El planteamiento consistió en obtener la observación d_0 , y realizar el cómputo de p , bajo el supuesto de que es cierta H_0 basándonos en los supuestos probabilísticos que procedan para poder hacerlo.

Este cómputo de p es el mismo de Fisher, pero se modifica el empleo que ha de dársele. Ahora se trata de apoyarse en él para adoptar una decisión. Este es un planteamiento singular, radicalmente diferente y opuesto en cierto sentido al de Fisher. Aquí aparece el propósito de adoptar una decisión, como todos conocemos, de este modo:

si $p < \alpha$, se rechaza H_0 a favor de H_1 ;
de lo contrario, se acepta H_0

Y finalmente, de una manera anónima e inercial, cerca del inicio de la II Guerra Mundial se unieron ambas teorías, se computa p y se empieza a decir que «la diferencia es significativa al nivel p ». En ese contexto se evita hablar de «aceptar» la hipótesis nula y los estadísticos se curan en salud diciendo que la rechazan o que no la rechazan.

Esta combinación de las dos teorías disgustaría tanto a los creadores de una corriente como a los de la otra. Ni Fisher por una parte, ni Neyman y Pearson por otra, congeniarían con la práctica contemporánea, porque ninguna de las dos escuelas estaba planteando nada como lo que actualmente se hace cada día.

Es especialmente elocuente el siguiente texto, escrito por los propios creadores de las pruebas de hipótesis: «ninguna prueba basada en la teoría de la probabilidad puede por sí misma generar índices válidos sobre la verdad o la falsedad de una hipótesis, las pruebas de hipótesis deben ser miradas desde otra perspectiva, siguiendo la regla de aceptar o rechazar una

hipótesis no estamos diciendo nada definitivo sobre si la hipótesis es o no verdadera, lo que se puede demostrar, es que si somos consecuentes con esa regla, a la larga la rechazaremos cuando sea cierta no más, digamos, que una de cada 100 veces si el «que se toma es del 1%, adicionalmente le rechazaremos con alta frecuencia cuando sea falsa»¹¹.

Esto, traducido a un lenguaje más familiar, sería similar a un sistema de justicia al cual no le concierne si un individuo en concreto es inocente o culpable, sólo trata de limitar el número total de veredictos incorrectos. Este es un rasgo loable de tal sistema, pero nuestro sentido de la justicia, como nos recuerda Goodman⁴, demanda que cada persona específica sea juzgada correctamente. Nuestra intuición científica nos dice, análogamente, que debemos tratar de obtener conclusiones adecuadas para cada estudio individual, y no sólo velar por conseguir que el porcentaje de veces en que nos equivocamos sea bajo.

Quería, para terminar con esta valoración crítica, comentar rápidamente qué efectos ha tenido todo esto en la literatura médica de hoy.

Por una parte, el tránsito de la evidencia empírica a las conclusiones es bastante formal. Por poner un ejemplo, al redactar un artículo científico, típicamente dará lo mismo si se obtiene $p = 0,04$ o $p = 0,001$; eso no cambiará los términos que elija el autor para exponer o comentar sus resultados. En ese mismo contexto quiero subrayar algo que para muchos ha de sonar muy familiar: el hecho de que sean frecuentes expresiones tales como: «puesto que $p = 0,12$, no hay evidencia muestral de que la hipótesis nula sea falsa». Esto escribirá el autor si a él no le gusta mucho la hipótesis alternativa. Ahora bien, si él estaba muy esperanzado en poder aceptarla, que es lo más frecuente, escribirá: «aunque $p = 0,12$ y no podemos decir formalmente que hay significación, con otro tamaño de muestra probablemente la habríamos encontrado». Y una vez más debo decir que esto me parece completamente legítimo, y comprendo el sentimiento de esta persona. En realidad, está diciendo «he seguido unas reglas de juego en las que no creo demasiado pero son las que se espera que yo use; pero estoy obteniendo un resultado que refleja una diferencia apreciable, ¿por qué tengo que renunciar totalmente a este indicio?»; es una manera de reaccionar como la de un «mal perdedor». Y digo que es legítimo, porque no es un invento: si obtuvo $p = 0,12$ ello se debe a que existe alguna diferencia, ¿por qué conducirse exactamente igual que si hubiera obtenido $p = 0,76$?

El otro problema es que la plausibilidad del resultado, los resultados procedentes de estudios previos y la solidez de la teoría no participan del proceso formal de la inferencia, sólo lo hacen de una manera difusa y, en cualquier caso, abierta a un considerable margen de subjetividad. Es un hecho objetivo, formalmente no

participan. Después puede uno aderezar, elegir palabras, etc.... en función de estos otros elementos a la hora de sacar conclusiones, pero lo cierto es que en el modelo formal no participan, y con relación a esto quería hacer una última cita, también bastante vieja, de Rozeboom, que plantea: «a pesar de la preeminencia que ha alcanzado este método, su empleo se basa en una incomprensión clave de la naturaleza de la inferencia racional; por ello las pruebas de significación casi nunca constituyen un medio apropiado para la investigación científica, si es que lo son en algún caso. Una hipótesis no es algo como un pedazo de tarta que uno puede aceptar o rechazar a través de una decisión voluntaria. Si hay algo que señala la irrelevancia práctica de las pruebas de hipótesis convencionales, es su incapacidad para proveer de genuinos recursos al investigador para su conducta inferencial»¹².

Luego de todo este planteamiento parecería como que nos hubiéramos enterado finalmente de que el Rey está desnudo: alguien lo dijo. Algunos colegas reaccionan airadamente contra este tipo de puntos de vista, porque sienten que lo dejan huérfano de recursos (este es un planteamiento que hago en mi libro) y arguyen que sólo aportan un enfoque destructivo sin ofrecer alternativas eficientes. A mi juicio, ésta puede ser una reacción bastante necia. Ponía en mi libro el ejemplo en el que estoy con una persona a la deriva en una balsa en el océano; ambos tenemos mucha sed y esa persona decide ingerir agua de mar; yo le explico que no la tome, porque hay un principio bioquímico según el cual se va a deshidratar más rápido y moriría antes. Le explico cuál es ese principio, pero aunque el individuo entiende que efectivamente no debe tomar esa agua, reaccionaría con necedad si dijera «está bien todo eso, pero no me valen tus argumentos porque yo tengo sed y tú no me dices dónde hay agua potable; voy a tomar agua salada». Afortunadamente sabemos donde está el agua potable: en el enfoque bayesiano.

¿Qué es, en esencia, el enfoque bayesiano? Sin entrar en tecnicismos, que no proceden dados los minutos de que disponemos, ha de consignarse que parte de un componente subjetivo; es decir, el investigador que está planteándose una hipótesis, comunica *a priori* su grado de convicción acerca de la validez de la hipótesis o, lo que es lo mismo, su grado de confianza en esa hipótesis y la medida en que no confía en ella. Eso lo plantea en términos probabilísticos, típicamente en términos de una distribución de probabilidades. Este componente es subjetivo, pero eso no quiere decir que sea caprichoso, son dos cosas distintas. Ha habido una tendencia a considerar como sinónimos a esos dos términos. Cuando se está dando por sentado que si yo soy subjetivo en un aspecto estoy siendo caprichoso, se está restringiendo mi derecho, y mi deber incluso, de hacer gravitar mi impronta per-

sonal a la hora de analizar la realidad, porque para algo no soy una Pentium III: soy un individuo con un cerebro con el cual voy a tratar de mirar críticamente la realidad y pronunciarme de alguna manera sobre ella.

Ya después hay un componente, que depende exclusivamente de los datos, la llamada «verosimilitud», y finalmente una conjunción de estos dos elementos, que produce una probabilidad final o *a posteriori* de la veracidad de la hipótesis. Dicho de otro modo: nosotros tenemos inexorablemente una visión *a priori* de la realidad. No podemos no tenerla, aunque los frecuentistas la olviden cuando hacen una prueba de hipótesis porque parten de una supuesta orfandad de criterios sobre lo que están examinando, como si no supieran nada y no tuvieran ninguna opinión. Por lo menos nada de esa percepción apriorística está dentro del modelo que emplean. Luego aparece esta realidad, que se mide objetivamente y que viene a modificar el grado de convicción que tenemos. Esto es consistente con lo que yo entiendo que es el elemento distintivo de la Escuela Bayesiana: parte de admitir la probabilidad no como una medición susceptible de ser estimada sólo a través de experiencias que permitan medir la frecuencia con que ocurren determinados acontecimientos, sino como un grado de convicción personal acerca de algo. Un grado de convicción que no depende de unos experimentos formales que se hayan realizado antes. Tanto es así que nosotros con frecuencia formulamos y contestamos preguntas tales como «¿cuán probable es que el Partido X vuelva a ganar las elecciones?». Es una pregunta razonable a la que se suele responder sin decir: «para responderla necesito que se produzcan 500 elecciones similares y contar cuántas veces ganó el Partido X».

Entonces, entre los rasgos claves de toda modelación Bayesiana hay que subrayar que no parte de un presunto vacío informacional sino que permite incorporar evidencias de las experiencias, de los experimentos y datos previos dentro de las conclusiones y dentro, incluso, del planteamiento inicial. Este enfoque, por otra parte, permite de manera natural y directa calcular probabilidades de eventos relacionados con observaciones futuras, algo que no puede hacer el frecuentismo; tal posibilidad es obviamente atractiva a la hora de tomar decisiones.

Otro problema (más técnico y complejo de explicar pero no menos real) es que las mediciones del frecuentismo dependen del diseño empleado y no sólo de los datos a los que su aplicación ha dado lugar: el cómputo del valor *p* depende de ese diseño, no sólo de los datos. Esto se debe al hecho de que para calcular dicho valor hay que incorporar información acerca de los valores posibles más extremos que el que objetivamente se observó. Se trata de valores que «podrían haberse observado» y esta potencialidad involucra al diseño que

se haya concebido. En dependencia de que haya sido uno u otro, los resultados posibles serán a su vez unos u otros. Por esa razón, en concordancia con esa lógica, una vez que se ha decidido un diseño dado, éste debe cumplimentarse tal y como se previó; no puede, por poner un ejemplo, incrementarse el tamaño muestral en medio de la experiencia si los datos que arrojó un primer intento experimental no resultaran suficientemente elocuentes (*modus operandi* del llamado *interim analysis*, cotidianamente empleado en los ensayos clínicos dentro del marco bayesiano: se continúa la actualización de los criterios que tengamos siempre que haya nuevos datos y el tamaño muestral no ha de establecerse antes). Así, un experimento puede detenerse en cualquier momento y por cualquier razón. Si se planifica una experiencia, un ensayo clínico, con el enfoque bayesiano, como no se ha de calcular ninguna p , el análisis no depende de la estructura del diseño, ni lo que se haya planificado de antemano, eso se puede explicar —como se ha dicho— con ejemplos que sobrecorren el marco de este debate. Esta flexibilidad es una de las razones que hace que hoy la estadística Bayesiana haya encontrado en los ensayos clínicos el campo de aplicación más fértil en los últimos 5 años. En general todo el proceso intelectual asociado a la inferencia bayesiana es mucho más coherente que los recursos clásicos con el modo usual que emplea el científico para pensar.

Quería terminar esta intervención inicial llamando la atención sobre un editorial de *Annals of Internal Medicine* del 15 de junio de 1999. En este momento creo que es la cuarta revista con mayor impacto del mundo en nuestro ámbito. Pues bien, entre otras cosas, se escribe lo siguiente (decidí traerlo en inglés, simplemente lo pongo tal y como allí aparece): «*Thus, just as clinicians need to know the likelihood that a particular patient has a disease given a certain test result, researchers (and those who read papers describing research) need to know the likelihood that a hypothesis is true given the data actually obtained in a particular trial or experiment*»¹³.

Estamos ante una teoría y ante una práctica emergentes. Ese editorial dice también explícitamente que *Annals* va a mirar con simpatía (ya lo hace desde 1997) cualquier trabajo realizado bajo el enfoque bayesiano, que este editor y esta revista consideran mucho más promisorio en cuanto a que podamos avanzar en el conocimiento científico.

Álvaro Muñoz

Bueno, yo también quiero agradecer la oportunidad de participar en este Congreso y en este Debate del cual ha debido aparecer el siguiente anuncio en El

Correo Gallego

Y las mayúsculas no son coincidencia.

Realmente, como Luis Carlos decía, el hecho de que el tema del debate no sea nuevo indica su dificultad, y lo que voy a tratar de mostrar es que existen problemas de fondo en relación a cómo se concibe por los bayesianos el proceso del conocimiento científico. Ya Luis Carlos elocuentemente indicó cómo la aproximación Bayesiana tiene el gran aspecto positivo de incorporar el estado del conocimiento de una manera explícita en el proceso inferencial mientras que para el frecuentista, los datos son datos que no se deben empañar con precedentes (*priors*). Luis Carlos indicó muy bien cómo lo subjetivo en el bayesianismo no se debe identificar como caprichoso y subrayó la valoración positiva de la capacidad del experto en influir sobre las inferencias. Desgraciadamente, en esta metodología, uno nunca está seguro del grado de capricho introducido por el experto, aun sin dudar de su buena intención. Este problema no lo tiene el frecuentismo y constituye una ventaja fundamental de este método: los datos no se despilfarran enderezando inexpertos.

El objetivo primordial de la epidemiología es diseñar estudios que obtengan datos para con ellos hacer inferencias y así reducir la ignorancia que tenemos sobre fenómenos de salud. Al entenderlos, racionalmente se pueden poner en marcha pautas y programas para mantener y mejorar la salud pública. La explicitación de inferencias tiene dos componentes fundamentales: i) cómo estimamos la naturaleza de la enfermedad que de una manera muy simplista, es lo que trata de hacer la media; y ii) cuál es el nivel de incertidumbre que tenemos con respecto a tal estimación. Es en la explicitación de la incertidumbre donde tal vez las dos metodologías se enfrentan de una manera más directa.

El manejo y la presentación del nivel de incertidumbre es el aspecto más difícil de lo que hacemos y es el que nos hace vulnerables a críticas vacías. No es difícil comunicar el significado de promedios y porcentajes, pero no es fácil comunicar medidas de incertidumbre evitando descalificaciones inmeditadas de lectores deterministas. En este sentido la aproximación Bayesiana tiene una gran ventaja sobre la aproximación frecuentista, porque la primera explicita automáticamente la incertidumbre que se tiene sobre un problema. En cambio la frecuentista necesita otro componente, es decir, necesita una varianza o un intervalo de confianza, y esta separación entre dónde está la inferencia y cuál es la incertidumbre es una desventaja del método frecuentista con respecto a la unión automática entre la media y la varianza en una aproximación Bayesiana. Al hablar así del bayesianismo debo aclarar: Yo no soy Luis Carlos, yo soy Álvaro.

Las dos aproximaciones se diferencian con respecto

a cómo se conceptualiza el progreso de la ciencia. La aproximación Bayesiana considera el progreso como un proceso continuo. Ya Luis Carlos decía cómo la posición bayesiana se parece o mimetiza todo el proceso de decisión que hacemos en la vida diaria; y es a través de esta posición que *Annals of Internal Medicine* se ha embarcado en la publicación de análisis bayesianos porque le está diciendo a sus lectores: «Miren, este proceso es cómo el clínico toma las decisiones.» Pero esto presenta muchos problemas. Aparenta ser muy benigno y adecuado pero asume que el proceso de la ciencia es continuo y acumulado y esto es, en mi concepto, *una filosofía de zapatero*. Realmente si uno lee a los verdaderos filósofos, no a los estadísticos, se da cuenta que estamos ante un tema de profundo debate. El progreso de la ciencia se produce a base de rupturas, de saltos, de cambiar de paradigmas. Los grandes avances en la ciencia, empezando por las aportaciones del maestro Galileo y todos los grandes científicos, han roto con los modelos aceptados hasta un momento dado. La manera continua, acumulada y tan amistosa de los métodos bayesianos dificulta la incorporación de cambios en paradigmas. En oposición, la aproximación frecuentista es más receptiva a rupturas.

Este es todo mi CUENTO sobre principios filosóficos, hasta aquí llego. De aquí en adelante le daré un giro radical a mi exposición. De aquí en adelante hablaré de la dificultad en la vida real, porque la estadística no es únicamente útil para hacer inferencias científicas; la estadística tiene otra gran función: la de estandarizar, regularizar y ayudar a establecer políticas y toma de decisiones. Es en este campo donde la aproximación Bayesiana tiene más problemas para establecer consenso científico. El bayesianismo se enreda con los precedentes; en cambio, en la aproximación frecuentista, el «parroquialismo» de utilizar solamente los datos facilita la consecución de consenso. El bayesianismo es vulnerable a intereses porque los intereses se pueden introducir en los análisis a través de los *priors*; en cambio, el análisis en la aproximación frecuentista se reduce a la «parroquia» de los datos.

Con respecto al establecimiento de regulaciones y políticas por agencias oficiales, en la aproximación bayesiana es difícil eliminar intereses no científicos y es vulnerable a pretenciosos y maliciosos; y no utilizo aquí la palabra caprichoso. Son los pretenciosos los más peligrosos, porque la ignorancia no lo es tanto. Lo que es más terrible es un poquito de conocimiento, ya que le da fuerzas a una persona para llegar a pretender y expresar su pretensión a través de los *priors*. Manejar a los ignorantes no es tan difícil, pero manejar a los pretenciosos es muy complicado. La actividad frecuentista facilita la acción regulatoria aunque, por otro lado, no está ni mucho menos libre de problemas, ya que en el frecuentismo, las políticas están al vaivén de estudios particulares.

En cuanto a su implementación se refiere, ambos métodos son inconsistentes. Es muy triste leer los artículos que proponen métodos bayesianos porque, por lo menos en mi lectura, de 10 artículos que uno lee, 9 utilizan precedentes planos (*flat priors*). Nadie se atreve, nadie se arriesga, a no poner los precedentes planos por miedo a influir y llegar a ser vulnerables a que se juzgue que las conclusiones son debidas a lo que se asumió en un precedente no-plano. Sin embargo, la gran ventaja de la aproximación bayesiana sería precisamente no hacer precedentes planos porque la incorporación del conocimiento no es plana. Por otro lado, también hay inconsistencias entre los frecuentistas. Ellos cortejan la aproximación bayesiana con todo el uso de modelos jerárquicos y de los modelos de variables latentes; también cuando utilizan, por ejemplo, los valores p unilaterales, pues su objetivo es tratar de incorporar conocimientos previos.

Para llevar a cabo análisis hoy día se requiere la disponibilidad de *software*. Afortunadamente para la aproximación bayesiana, en este momento, se empieza a disponer de *software*, lo cual hace la metodología factible. Una propiedad atractiva de la aproximación bayesiana es que requiere conocimiento para poder hacer el análisis y esto es muy importante. En la frecuentista, la gran cantidad de *software* disponible es su propio enemigo, porque hay un gran abuso por parte del novato en la utilización de estas metodologías. El abuso de la regresión de Cox en la última década es un ejemplo canónico. El enfoque bayesiano requiere pensar y, en ese sentido, evita automáticamente los abusos.

Otro aspecto fundamental de nuestro trabajo profesional y académico es la comunicación de nuestros resultados al público en general. Esto es esencial si queremos influir en las políticas de salud. Es más, cuando comunicamos, necesitamos incluir métodos efectivos para educar al público sobre lo que hacemos. Una gran ventaja de la aproximación bayesiana es que la incertidumbre es tan importante como la señal del mensaje. En contraste, usualmente los resultados frecuentistas se presentan de tal forma que la varianza aparece en letra menuda y la media en letra grande. Esto no es adecuado, y argumentos de simplicidad en las presentaciones terminan por engañar al público y perpetúan el rechazo al ruido o incertidumbre que necesitamos predicar.

¿Cómo educamos a nuestros alumnos? ¿Qué hacen las Escuelas de Salud Pública? Este es un gran dilema. El verdadero bayesiano dedica su tiempo a tener buen conocimiento, buenos precedentes, y a aprender sobre el **tema**. Pero esto se hace a costa de no dedicar recursos y esfuerzos para aprender sobre los **métodos**. El frecuentista, por otro lado, se concentra en los métodos a costa de una desvinculación con el conocimiento que le permitiría tener buenos precedentes. Es un problema de tiempo.

Desde esta perspectiva, la aproximación bayesiana mantiene la epidemiología como eje de investigación en salud. La base de la epidemiología se puede describir por la terna: BIOLOGÍA, DATOS, métodos (donde métodos representa métodos para *odds ratios*: la bandera de las medidas de asociación de epidemiología). En cambio, el frecuentista necesita la estadística como el eje de investigación en salud porque en esta aproximación la base corresponde a la terna: biología, DATOS, MÉTODOS (no sólo para el *odds*). Los DATOS son comunes a ambas aproximaciones y disciplinas. Yo tenía cierta duda en hacer estas afirmaciones porque, como esto es un Congreso de Epidemiología, ustedes concluirán que de aquí en adelante todos ustedes van a ser Bayesianos. Este no es mi mensaje. La epidemiología va mucho más allá de los métodos. Es la piedra angular para describir fenómenos de salud y establecer políticas que conserven y mejoren la salud pública. Los debates metodológicos deben mirarse desde esta perspectiva, y nunca perderla de vista.

Moderador

Luis Carlos dispone de 10 minutos para rebatir.

Luis Carlos Silva

Voy a hacer una pequeña confesión, una infidencia, al público. Álvaro y yo tuvimos tres contactos antes de esta reunión. Ese es un dato que en principio obra contra la lozanía del debate; pero debo aclarar que esos tres contactos duraron unos 15 minutos cada uno: el primero fue hace un año en Barajas, unos minutos antes de regresar él a Estados Unidos; el segundo fue en mayo de este año en Baltimore y duró el tiempo que nos llevó comernos un emparedado en la cafetería de la Escuela de Salud Pública de Hopkins; y el tercero fue ayer, almorzando con Jordi Sunyer, aquí en Santiago de Compostela, de modo que hay poco amañamiento. No obstante, sabíamos más o menos qué íbamos a decir cada uno en esta primera intervención; lo que no tenemos preparado y cada uno de nosotros ignora es lo que el otro ha de decir en este turno de réplica.

Cuando Álvaro ponía ahí «el cubano bayesiano» (se me ocurría que también pudiera haber puesto el «cubayesiano»); sin embargo, pensaba que debo aclarar que a mí lo que menos me gusta de la teoría bayesiana es el nombre, porque éste evoca a una secta. Preferiría que alguien lo bautice de otro modo, quizás «estadística racionalista» o algo parecido, porque el enfoque bayesiano, en su espíritu, es exactamente la an-

títesis del pensamiento sectario. Dicho esto, quiero hacer algunos comentarios sobre las críticas que él hacía.

Hay un elemento de subjetividad indiscutible a la hora de identificar los *priors*, si es que realmente se hace tal elección en lugar de optar por lo más cómodo, lo más perezoso, que sería adoptar los *priors* planos, ininformativos. Pues bien, es cierto que hay un elemento de subjetividad; lo que pasa es que el bayesianismo asume esa subjetividad, acepta que no se puede no ser en uno u otro grado subjetivo; el frecuentismo, en mi opinión, no se libra de ella, sino que la esconde (o la disfraza, o la suple), con esquemas o decisiones difícilmente explicables. Por ejemplo, consideremos el alfa 0,05, para detenernos en los guarismos sacralizados. Como algunos me han oído decir en los cursos, hasta ahora no he encontrado ninguna explicación más persuasiva de que se haya escogido 5 y no otro número que el hecho de que tenemos 5 dedos; y eso sí, está bien, es más objetivo que un elemento subjetivo, pero en cierto sentido es más caprichoso.

Segundo, en cuanto a que el bayesianismo es resistente a las rupturas, también es cierto; si de improviso aparece alguien con ideas muy novedosas y heterodoxas, el bayesianismo se resguarda de ello, porque le dará un grado inicial de credibilidad bajo, de manera tal que, cuando los datos se conjuguen con el grado de credibilidad *a priori*, el nivel de credibilidad *a posteriori* será reducido. Es resistente a las rupturas, pero a mí no me inquieta mucho eso; yo veo tantos iluminados por ahí, tanta gente planteando cosas innovadoras, sin ton ni son (no voy a mencionar a nadie para no herir ninguna sensibilidad), pero hay tantos terapeutas alternativos... que la verdad es que tengo ganas de tratarlos bayesianamente cuando vienen con un experimento que ha dado un resultado «significativo» (no hablemos ya de los que se basan sólo en anécdotas espectaculares). Si usted es innovador, pues «cúrrselo», pase trabajo para demostrármelo, exponga muchos datos, y si los datos tienen mucha elocuencia en la dirección de lo que afirma, entonces ya las probabilidades *a posteriori* serán un poco más altas de lo que eran originalmente, quizá no mucho, pero la siguiente vez, ya se tendrá menos derecho a aceptar una probabilidad *a priori* tan baja, porque ya existirá un estado de conocimiento orientado de alguna manera a favor de ese planteamiento novedoso. Si por el contrario, sus datos no son tan convincentes, entonces ese «fuego artificial» no será confundido con la pólvora.

Si alguien elige, por razones de comodidad, los precedentes planos, no creo que sea un problema del enfoque bayesiano, sino un problema del usuario. Si él se siente poco a cubierto, si siente poca confianza en su propio sentido común y en la cultura que posee sobre el asunto estudiado, eso es un problema de quien está utilizando esa herramienta con tal timidez; pero en definitiva está bien que se conduzca así en ese caso. Cabe

recordar la regla de oro de la ciencia: máxima audacia en la formulación de hipótesis y máxima cautela al sacar conclusiones. Entonces Ud. tiene que tener audacia, sin audacia no se avanza; ya el teorema de Bayes se ocupará de reducir el optimismo si los datos no lo refrendan. Entonces creo que unos precedentes elegidos adecuadamente son una expresión de audacia que a mi juicio sería muy bienvenida en el mundo contemporáneo.

En cuanto a la imputación de que el método bayesiano es vulnerable a los intereses, he de decir que también es cierto. Sin duda alguien, a través de los *priors*, puede de antemano manipular de alguna manera, en función de ciertos intereses, los resultados de los estudios. Ahora bien, el frecuentismo no está en mi opinión exento de ese riesgo, porque ¿qué cosa es el resultado de una prueba de significación? Números, no conocimientos. Los números son, al final, números, expérense en términos de tasas de error, probabilidad de haber cometido el error de primer tipo o simplemente en forma de una tasa o porcentaje de individuos que cumplen una condición. Pero esos números son intermediarios metodológicos para una pregunta que nunca se responde sólo con ellos. Si tengo que dar una respuesta sustantiva al problema, sea yo frecuentista o bayesiano, si quiero trascender, tengo de traducir esos números, esos resultados, en conclusiones sustantivas; para eso tengo que elegir palabras, tengo que elegir adjetivos y tengo, quizás (no el compromiso de sugerir soluciones, porque no es realmente la tarea del estadístico como tal) que cerrar eso en algo más allá que los números. Porque hasta ahí llega el SPSS; a partir de eso empiezo yo. Y eso es así, sea bayesiano o no sea bayesiano, y en esa interpretación, en esa traducción, también cabe, como todos sabemos, un posible componente donde pudieran estar involucrados intereses. Por otra parte, si uno manipula conscientemente los *priors* en función de ciertos intereses, entonces se trataría de un acto de deshonestidad, y eso no es un asunto del enfoque bayesiano; eso es una felonía intelectual, que pudiera cometer cualquier inescrupuloso (alguien podría por ejemplo inventar datos, sin más, por muy frecuentista que sea).

Quería terminar esta intervención haciendo una pequeña crítica. Hoy estuve por la mañana en la mesa que trató los temas de las estimaciones en áreas pequeñas. Fue muy interesante, me sentí muy satisfecho allí, puesto que identifiqué que hay un grupo, muchos de cuyos integrantes no se conocen entre sí, como pude corroborar, pero un grupo bastante sólido de gente trabajando en esta línea, aplicando métodos bayesianos. Sin embargo, a todos les quiero hacer una observación que no quise hacer allí porque me parece más oportuno este otro espacio. Y es lo siguiente: todos empezaron su trabajo explicando las dificultades que tenía el enfoque frecuentista para abordar el problema que estaban tratando de resolver: denominadores demasiado

pequeños en determinadas áreas, consecuentemente tasas muy lábiles. Entonces el enfoque bayesiano era la alternativa que utilizaban. Magnífico, pero es que lo están diciendo con un sentimiento de culpa; es decir: «perdóneme por usar el enfoque bayesiano, pero es que el otro no me sirve». Y a mí me parece muy bien este enfoque porque es en sí mismo interesante, en sí mismo reproductor de un modelo de pensamiento que yo vuelvo a insistir que me parece más racional y en eso, desde luego, Álvaro estaba de acuerdo conmigo desde el principio, y lo expresó aquí.

Para terminar quiero hacer una pregunta a Álvaro, ya que le toca usar la palabra en los próximos diez minutos. Quiero saber si realmente tú crees, Álvaro, que el siglo veintiuno no será necesariamente un siglo bayesiano.

Álvaro Muñoz

La probabilidad de que el siglo veintiuno sea bayesiano es de 0,23, solía ser 0,15 hace cinco años, pero luego de los editoriales en el *Annals of Internal Medicine* ha aumentado a 0,23. Quiero solamente reiterar que estos comentarios que yo hice son influenciados por la enseñanza que recibí en estadística matemática y por la práctica que he tenido en salud pública desde 1980, que es muy cercano al año en que nací.

En respuesta a Luis Carlos, quería hacer dos comentarios adicionales. El primero es que la crítica de los valores p no es una crítica a la estadística y a sus principios. Los valores p y el culto al 5% han sido hechos por los reguladores, porque se necesita un número «mágico» para poder regular. Si uno mira el valor p como otra forma de cuantificar la incertidumbre que uno tiene, se le elimina el culto al «nivel de significación del 5%».

El segundo punto, que tal vez es el gran problema del bayesianismo, es la manera tan simplista de incorporar el conocimiento. La manera de incorporar el conocimiento y hacer consenso científico es mucho más complicada que el teorema de Bayes. Aunque tiene desde el punto de vista probabilista toda su connotación, no es así como progresa la ciencia. La ciencia es muy extraña, es como tierra pantanosa. Y solamente por formas muy inusitadas las cosas van surgiendo y se identifica cuál es la verdad; tal vez la persona que en epidemiología ha escrito sobre el tema en forma más elocuente y provocadora es el holandés Vandenbrouke¹⁴. Él sugiere que la forma cómo entendemos y describimos las cosas al final no es un estudio, no es un *posterior*, es una cantidad de información que se va propecolando y destilando. En este contexto la secuencia de reacciones cuando leemos la literatura sobre un tema sigue los siguientes pasos: al primer artículo: esto no

es verdad; al segundo: sí es verdad pero no tiene aplicaciones prácticas; al tercero: sí es verdad y tiene aplicaciones prácticas, pero ya todos lo sabíamos.

Debate

Moderador: Aunque podríamos dejar que los ponentes sigan, es hora de abrir el debate al público, por lo que os invito a que hagáis las preguntas que consideréis... Me pregunta Luis Carlos que si mientras os animáis no tengo una pregunta, le contesto que sí:

La pregunta tiene algo que ver con la anterior referencia al siglo veintiuno. Bacallao⁵, en el *Medicina Clínica* del 96 cita a Efron diciendo que al enfoque clásico le augura una larga vida y al bayesiano una larga espera. Le pregunto a ambos, ¿qué opináis de ese juicio ahora, trece años después?

Luis Carlos: Desde luego, Efron escribía esto hace trece años¹⁵, y habría que ver a qué le está llamando «largo». Lo que antes era largo, ahora es larguísimo; los lapsos se acortan, el mundo se achica, todo es más rápido, y las expectativas son superadas por las posibilidades de multiplicación de uso de tecnologías, de intercambio de opiniones e información, y por la posibilidad de comunicación horizontal que ofrece Internet, que para mí es uno de los rasgos más interesantes de este recurso, la cual está también contribuyendo a que determinadas dictaduras metodológicas se tambaleen.

Pienso que el arribo del pensamiento bayesiano, por otra parte, nos coloca ante un enfoque opuesto al convencional. Que nadie se asuste: la estadística que se aprendió no es una estadística inútil para hacer un análisis bayesiano. Si usted aprendió las técnicas de regresión lineal, ellas siguen siendo útiles aunque el enfoque sea bayesiano, porque esa parte sigue incólume. Pero, volviendo a lo que decía, en cierto sentido ya ha sido una larga espera, porque trece años, en esta época, es una espera larguísima. Ya ha transcurrido el tiempo y lo que cabe pensar, lo que todos vemos, de hecho, palpablemente, no sólo por un editorial aislado de *Annals*, es una irrupción sensible de estos métodos. Están los datos que el propio Enrique dio cuando empezamos, y todos tenemos ya acceso a un *software* bayesiano, un hecho que hace cinco años era impensable. Y pienso que en otros cinco años o poco más todos los grandes paquetes comerciales tendrán que incluirlos.

Tampoco creo que cometa una infidencia muy importante diciendo que nosotros estamos trabajando en la nueva versión de EPIDAT, que es un paquete estadístico en el que han venido trabajando colegas de Galicia y de la OPS. En la versión 3 para los entornos actuales de Windows va a incorporarse un módulo bayesiano. Creo que, pese a que en aquel momento tenía que soportar una larga espera, ésta es ahora mucho

más corta. No sé si alguna vez, por poner otro ejemplo, en algún Congreso anterior de esta Sociedad se había producido aunque sólo fuera un único trabajo que tuviera enfoque bayesiano, o dos o tres; hoy, tuvimos aquí, en este Congreso, unos cuantos.

Álvaro: Yo voy hacer un comentario histórico. Realmente el énfasis o la aparición de la aproximación bayesiana es un resurgimiento, porque en el siglo pasado, en la época de Bayes y de Laplace, todo el mundo era bayesiano; inclusive Gauss cuando propuso la teoría normal. Él mismo, que se le presenta ahora como frecuentista, al principio era bayesiano¹⁵. Su posterior adscripción al frecuentismo se debió a Fisher, porque éste no trabajó únicamente en áreas inferenciales de la estadística sino también en cómo se enumera en estadística y cómo se diseñan estudios. Por lo tanto, no hay que dejarse impresionar con que el bayesianismo está surgiendo solamente ahora. No hay tal; es, desde el punto de vista histórico, un resurgimiento.

Marina Pollán: Yo quería aclararle a Luis Carlos, porque estábamos en la misma mesa, que en principio disculpe que la gente presente así los resultados, pero todos primero aprendimos a pensar como frecuentistas siendo médicos, lo que nos supuso un esfuerzo importante. Muchos de nosotros seguimos sin entender cabalmente qué es un intervalo de confianza, y los que lo entendemos de vez de cuando lo olvidamos. Por ello creo que, honestamente, la intención de las personas que estaban allí era mostrar cómo se vieron forzados a entrar en la dinámica bayesiana. Todas las presentaciones de esta mañana carecían de variables explicativas; por lo tanto los *priors* eran planos y tenían sentido, porque lo que se asume bajo este contexto son los *priors* centrados en la hipótesis nula. Cuando estás estandarizando con tu propia población, con todo el área geográfica, es lo esperado. Pero creo que el problema empieza cuando tratas de introducir variables explicativas. Yo sólo tengo algo de experiencia en modelos bayesianos en el ámbito del análisis geográfico. Cuando empiezas a introducir variables explicativas, no conozco ningún estudio que se haya mojado en utilizar un *prior* que no sea no informativo y ahí sí que existe conocimiento anterior, o siempre intentas buscar variables ecológicas que tengan cierto sentido bien porque haya literatura detrás o un pensamiento plausible que al menos te diga cuál es la dirección de la asociación.

Nosotros hicimos un estudio muy modesto cuando empezamos a aprender esta historia de los métodos bayesianos. Utilizamos un modelo bayesiano y un modelo jerárquico frecuentista. Tengo que decir que publicamos el frecuentista; nos sentíamos más cómodos hablando en el marco del frecuentismo, porque nos costó mucho aprenderlo. No sé si lo manejo bien, pero desde luego ya tengo cierta experiencia en ese marco mientras que en el otro ámbito te mueves con bastante más incertidumbre. Pero también tengo que decir, sincera-

mente, que a mí me parecía que en cuanto a este tipo de modelos jerárquicos, los resultados del modelo bayesiano mostraban más respeto hacia los datos. En los modelos geográficos es muy importante resaltar que una vez que incorporas variables explicativas, quieres saber qué pasa dentro, qué cosas no explica el modelo, qué extravariabilidad hay en tus áreas geográficas inexplicada con las variables que tienes. Porque esto a la vez te genera nuevas hipótesis. En este aspecto, creo que los modelos bayesianos son más respetuosos.

No sé si Álvaro tiene más experiencia, más conocimiento y puede decir algo más al respecto. Lo que realmente es un quebradero de cabeza brutal es el asunto de cuántos ciclos, cuál es el *burning* que tienes que utilizar, es decir, cuándo puedes pensar que el modelo ha convergido y que ya está en la zona adecuada de muestreo de estimadores. Otro quebradero de cabeza es saber cómo valoras tu modelo, cómo sabes que el modelo que tienes es razonablemente adecuado. Y ese es el caballo de batalla que yo creo esperamos que nos digáis por donde hay que seguir.

Yo tenía una pregunta concreta para vosotros: cuando tienes variables fijas y variables aleatorias, modelos jerárquicos por decirlo de alguna manera, según vuestra experiencia o vuestra intuición, ¿creéis que ambas aproximaciones frecuentista y bayesiana, funcionan igualmente bien?

Álvaro: Yo creo que Marina ha dado un ejemplo de que, desde el punto de vista filosófico hay una gran dicotomía, pero no en la práctica. Todos sabemos que cuando hacemos un análisis de datos no es que haya un modelo; existen muchos modelos para hacer las inferencias. De la misma manera, es posible que las metodologías que debatimos no sean tan opuestas. Realmente, creo que no debemos de perder de vista lo que tratamos de hacer: reducir nuestra ignorancia sobre los procesos de la enfermedad en las poblaciones. Esa es la meta. La gran ventaja del uso de la metodología bayesiana, que es un gran avance en el proceso de reducir la ignorancia, es tratar de incorporar conocimientos previos sobre las inferencias que uno quiere hacer en estos momentos. Pero no sólo conocimientos al nivel de una persona, sino una *prior* que capture el estado de conocimiento. Y así, decidir cómo esos datos, que se obtienen en un estudio específico, están de acuerdo, mejoran, o discrepan del estado del conocimiento que se tiene en un momento dado. Al hablar de esto, Vds estarán seguramente pensando que estoy reinventando la combinación bayesiano-objetivo, y esto ya está derrotado. Es por eso que me presenté como un cuentista más que como un frecuentista.

María José Tormo: Gracias, me ha gustado mucho la discusión. A esto le llamo aprender disfrutando y enhorabuena por conseguir el alto nivel de la sesión. La verdad es que, cuando ha empezado la sesión, no tenía

una idea muy clara sobre los dos métodos y, en cambio, ahora siento que he avanzado en su comprensión y me decanto, seguramente no importa mucho por lo que yo me decante pero lo voy explicar, por lo frecuentista. Y estas son mis razones:

A mí me han enseñado que cuando empiezas a andar y, sin querer, te alejas un poquito del sendero original, al cabo de 100 km. puedes estar tan desviado que ya no recuerdes cuál era el camino original. Creo que eso puede ser un riesgo grave de los bayesianos y sus valores previos de partida (*priors*). En cambio, el frecuentismo me recuerda a esa frase tan bonita que tenemos en cuenta siempre cuando investigamos «Vamos a echar una mirada fresca sobre un problema viejo», una mirada limpia, sin pre-juicios. Me parece que esto es más lo que es deseable en ciencia: partir de cero, tratar de romper el conservadurismo, romper los viejos esquemas. Plantearse la investigación como un nuevo descubrimiento.

Por otra parte, creo que en el manejo de las «*p*» del frecuentismo —cuyo abuso es una lacra— los epidemiólogos y los estadísticos han llegado afortunadamente a una cierta madurez que es la que se tiene cuando las *p* descansan debajo de nuestra voluntad y solamente las dejamos salir cuando es preciso, cuando aportan más luz.

Más aún, de lo que no hemos oído hablar nada es de que el bayesianismo, las técnicas bayesianas, hayan resuelto algún problema. Esta mañana hemos visto una sesión excelente sobre la representación gráfica de la enfermedad, pero me gustaría saber qué otros problemas candentes de la salud y la enfermedad han podido resolver los métodos bayesianos que no hayan podido avanzar los frecuentistas. De eso no habéis mostrado ningún ejemplo, y por tanto, yo me quedo con la necesidad de un poco de empirismo.

Y por último, recordar que, al menos en España, donde la mayoría de los que trabajamos en epidemiología somos de formación médica, debemos sentirnos cómodos con nuestra formación —es decir, de contar con muchos puntos de partida o *priors*: médicos, biológicos, patológicos, etc., sacarles el mayor provecho —sin complejos— y dejar que los estadísticos —preparados formalmente más para estas tareas tan sofisticadas o ¿incipientemente desarrolladas?— avancen en los métodos, los refinan y los hagan inteligibles siempre contando, muy estrechamente, con la colaboración de los epidemiólogos. Y que esta unión sea cada vez más fértil. Gracias.

Moderador: Creo que le toca a Luis Carlos.

Luis Carlos: Dos comentarios muy rápidos. Uno es que yo sinceramente no creo que el frecuentismo parta de cero; creo más: que eso es imposible. No es posible dar una mirada enteramente fresca a nada. Todos tenemos prejuicios, una visión, un compromiso, una convicción, y no podemos decir, a partir de las cinco voy

a dejar de pensar, me voy a olvidar de lo que ya sé, voy a hacer un ejercicio de amnesia. Ni para esto ni para sacar conclusiones. Al revés, lo que creo es que muchas veces en el fondo lo que está haciendo un individuo que dice que p fue 0,12 pero que con un tamaño de muestra mayor quizás hubiera obtenido otro resultado, es pensar bayesianamente por encima de los cánones de la elección que hizo, es revelarse instintivamente contra la obligación de dar una mirada «fresca» a lo que ha observado muchas veces.

Sobre los resultados que no se habían mencionado que pudiera haber producido el bayesianismo, me veo obligado a hacer una reflexión de tipo completamente conceptual y general. El bayesianismo no ha resuelto absolutamente nada; o sea, no ha descubierto absolutamente nada; pero el frecuentismo tampoco ha descubierto absolutamente nada, nunca. Es decir, no es el frecuentismo o el bayesianismo quien puede resolver problemas, los que resolvemos problemas somos los que nos planteamos los problemas. Tenemos unos recursos metodológicos y tecnológicos para usarlos en el proceso de tratar de resolver esos problemas. En ese sentido me parece una falsa pregunta como lo sería indagar acerca de cuáles son los problemas que han sido «resueltos por las pruebas de hipótesis» como tales. Las pruebas de hipótesis no han resuelto nunca ningún problema. Yo incluso reacciono con cierta irritación cuando un alumno pone en las conclusiones de su tesis que «el resultado de la aplicación de la prueba es significativo». Le explico que eso a esa altura no interesa; en ese punto lo que interesa es el partido que se le saca al hecho de que el SPSS me haya informado que la diferencia es significativa. Lo mismo le diría a un bayesiano: ¿qué partido interpretativo y conceptual le vas a sacar a esta nueva distribución de probabilidades que tienes ante ti?

En segundo lugar, creo que es un reclamo un poco injusto también por el hecho de que nosotros sólo hace cinco o seis años que contamos con la posibilidad factual de hacer análisis bayesianos. Aclaro: cuando digo «nosotros» me refiero a todos nosotros, no a «nosotros los bayesianos» ya que yo no lo soy como tal: simplemente simpatizo extraordinariamente con ese modo de pensar.

¿Con qué recursos se contaba? No había recursos computacionales simplemente, mientras que la prueba de Ji cuadrado se puede hacer con una calculadora, y hasta una regresión múltiple, aunque lleve un poco de tiempo, también se puede hacer con una calculadora, un estudio de Montercalo, con mil «pruebas de calentamiento», es simplemente imposible de realizar sin una computadora, No alcanza la vida de nadie para hacerlo. Ahora es que tenemos recursos que nos permiten que nos alcance una tarde, quizás, para ello. Esa pregunta, suponiendo que no fuera válida la objeción anterior, pues sería justo que se formulara avanzado

el siglo XXI.

Alvaro: Yo quiero decir algo. Ella no quiso decir eso. Ella, cuando habló de una mirada nueva a un problema, no quiso decir lo que tú pensaste que ella dijo. Porque apreciaba y valoraba el conocimiento que había adquirido como clínica; y, aun cuando iba a tener una mirada nueva, ella estaba incorporando de alguna manera el conocimiento que tenía. Entonces es injusto tomar las palabras literalmente, y más aún por un bayesiano.

Por otro lado, es fundamental incluir el conocimiento adquirido en la conducción del análisis. Si no es formalmente a través de una *prior* en la aproximación bayesiana, el investigador/analista debe guiar el análisis frecuentista de acuerdo al estado del conocimiento. Aún más importante, el estado del conocimiento debe ser incorporado en la etapa del diseño del estudio. En este contexto, quiero compartir con ustedes una anécdota. Recuerdo vivamente el último consejo de mi tutor (Bill Brown) en la Universidad de Stanford cuando terminé el doctorado: «Mire, hombre, Álvaro, cuando vaya a colaborar con alguien asegúrese de que sea buen investigador, que sepa tanto que no se necesiten datos».

Marisa Rebagliato: Bueno yo también quería sumarme a la felicitación de M.^a José, y supongo que también por la experiencia de enseñanza lúdica que me habéis hecho vivir o que nos habéis hecho vivir. Creo que hacía tiempo que no me había reído tanto en un Congreso de Epidemiología como en el día de hoy. Y bueno, mi pregunta era: como enseñantes que sois ambos, quería que diérais una predicción a corto plazo, en un futuro inmediato, sobre las posibilidades o retenciones que va a haber a la hora de implantar la formación o la enseñanza de este tipo de enfoque en lo que son Universidades, en formación de posgrado, porque claro, si queremos que haya debates de este tipo, que haya masa crítica, que haya estudios bayesianos con *priors no flats*, necesitamos formación. Y entonces os pregunto: ¿cómo veis el panorama?

Luis Carlos: Bueno, rápidamente, yo puedo decirte más que una predicción, puedo darte un testimonio, modesto, pero elocuente. En particular, ya dirigí actividades docentes de posgrado en cursillos y seminarios, impartiendo o conduciendo discusiones en torno a este asunto, que han sido muy exitosos y muy bien recibidas por médicos, por clínicos. Por lo menos, hay dos concretas: una fue aquí en Galicia (muchos de los que participaron en esa experiencia están presentes aquí); la otra en Asturias y se prevé que haya dos o tres más, pronto. Creo que es una pregunta muy pertinente, porque en definitiva, a mi juicio, está aludiendo a un problema muy serio: el planteamiento del enfoque bayesiano exige otra manera de pensar, que no es que sea radicalmente opuesta, ni que sea completamente novedosa, ni que exija empezar de cero; pero es otra manera de entender la realidad y otra ma-

nera de pensar, y ello produce un coeficiente de rozamiento alto. Por el momento se interesan grupos relativamente reducidos de colegas que tienen una vocación de reflexión epistemológica, y que están tratando de buscar respuestas, o sea, recursos alternativos. Aunque está claro que se trata de un proceso embrionario. Y ese era un poco el sentido que tenía este debate. Para la mayor parte de los colegas aquí presentes, me imagino que este ha sido un ejercicio de hacerles saber que hay un debate. Aspiro a que se hayan enterado de los elementos más básicos de éste; por lo menos saben que existe, cuando vean artículos publicados con estos enfoques no dirán «de qué rayos me están hablando». Mi visión es en cualquier caso, muy optimista.

Álvaro: yo estoy tan confundido como tú, Marisa.

Tomás Vega: Quiero agradecer en primer lugar vuestra conferencia-debate que, para los profanos en el tema, ha resultado sumamente ilustrativa. De toda la exposición, me ha sorprendido particularmente, que los criterios principales para decantarse por la metodología bayesiana o la frecuentista se refieren al tamaño de la «n» y a las tasas inestables, más que a la calidad de los datos.

Se me ocurre pensar en los estudios destinados a la toma de decisiones, en los que una actitud conservadora pudiera estar indicada y en consecuencia el rechazo de una hipótesis nula (con un modelo frecuentista) y la subsiguiente decisión pudiera entrañar un riesgo importante.

Me gustaría saber si existen circunstancias de este tipo que nos permitan inclinarnos por uno u otro método; ¿hay algún criterio que permita aumentar la probabilidad o verosimilitud de acertar con el método elegido? Gracias.

Luis Carlos: Hay una cosa que hace un poco difícil responder tu pregunta. Ha de quedar claro que el enfoque bayesiano nunca se plantea rechazar ni aceptar nada: hay una hipótesis que tiene un grado de credibilidad; según el *status quo*, se hace un examen y se actualiza ese grado de credibilidad. Justamente, una de las cosas que más se critica es la pseudo decisión que uno toma cuando hace una prueba de hipótesis. Bien, eso es lo que plantea teóricamente todo el enfoque frecuentista: que se actúa para tomar una decisión. Y en realidad esto no es verdad: todavía me falta por conocer el primer investigador que tras una experiencia que resultara no significativa haya desechado definitivamente la hipótesis en la que creía. La comparación no es enteramente posible porque se están planteando tareas estadísticas distintas. En el fondo no son tareas distintas, como bien señaló Álvaro cuando empezó, la tarea es la misma (tratar de reducir nuestra ignorancia, eso es lo más general), pero a la hora pragmática de actuar, las tareas estadísticas que se están planteando uno y otro enfoque son distintas, por lo menos teóricamen-

te. Sin embargo, como ya dije, creo que los frecuentistas, en el fondo, al final razonan bayesianamente, a pesar de su esquema, sólo que no lo reconocen, o no lo saben.

Santiago Pérez-Hoyos: Una pregunta a raíz de la última afirmación de Luis Carlos; dice que los frecuentistas razonan bayesianamente... y ¿por qué los bayesianos utilizan mil simulaciones y calculan una frecuencia para resolver sus problemas?

Luis Carlos: El hecho de que haya una filosofía, una conceptualización bayesiana general, no significa que se renuncie a ningún otro recurso que pueda ser eficiente. Cuando yo calificaba que un sistema de justicia que asegure que el 95% de sus veredictos sea correcto como algo loable, estaba elogiando un rasgo de la estadística frecuentista. Eso es una manera de decirte que no se hace una negación doctrinaria de los resultados de la estadística convencional o frecuentista, lo cual sería ridículo. Lo que se está planteando es, y eso sí es una mirada fresca, que determinados procedimientos ya huelen a algo bastante quemado; pero eso no quiere decir que se empiece de cero y que nada de lo anterior valga.

No identificado: Quería hacer un comentario y una pregunta. Una de las críticas que se le ha hecho al enfoque frecuentista, es que para qué se va a contrastar una hipótesis que nunca se va a aceptar. Realmente en un contraste de hipótesis, la hipótesis nula nunca se acepta: se rechaza o no se rechaza. En la práctica, el contraste de hipótesis se ocupa de valorar si tenemos suficientes argumentos para aceptar la alternativa, si es demasiado probable que nos equivoquemos en caso de aceptar la alternativa. Si no la aceptamos, eso no significa que aceptemos la nula.

En cuanto al procedimiento del bayesiano quiero hacerle una pregunta al Profesor Álvaro Muñoz: una de las críticas que hace frecuentemente al enfoque bayesiano es que establece *priors* no informativos, o trabaja con varianzas muy grandes, o de precisión muy baja. Entonces, en el fondo, cuando se establecen tales *priors*, el investigador no se quiere «mojar» sobre su grado de conocimiento, sobre el problema. Al actuar así, implícitamente, casi todo el peso para la obtención de las distribuciones a posteriori se traslada a la verosimilitud; de modo que la obtención de la distribución a posteriori se basa fundamentalmente en los datos empíricos y de hecho se le está dando razón a los frecuentistas. ¿Está Ud. de acuerdo?

Álvaro: Esta es la versión activista de los comentarios que he hecho sobre las *flat priors*. Utilizar *flat priors* es antónimo al objetivo de la aproximación bayesiana. En parte está relacionado con el sentimiento de culpa al que aludía Luis Carlos e influenciado por el temor del bayesiano de que se piense que las conclusiones son debidas a la utilización de *priors* que no sean no-informativos. Pero realmente, el gran poder de la aproximación bayesiana es no asu-

mir *flat priors*. A mí me gusta que usted haya utilizado la palabra «no se moja» cuando se refirió a las *priors* no-informativos. Recuerden aquella alegoría que yo les mencioné, es que la ciencia es una tierra muy pantanosa y realmente si se va a hacer ciencia, hay que mojarse.

Moderador: Para terminar, decir que para mí, en lo personal, ha sido un honor estar aquí entre Álvaro y Luis Carlos. No hace mucho tiempo escuché un comentario acerca de que los Congresos, con el paso del tiempo, los recordamos por la vertiente lúdica, por lo buena o mala que fue la comida, por la fiesta aquella que acabamos a las cinco de la mañana.... Esa es una opinión

con la que no estoy en total desacuerdo, pero de lo que sí estoy seguro es de que este congreso lo recordaremos dentro de unos años por la fiesta de mañana y por la sesión que ahora concluimos. Gracias a todos.

Bibliografía

1. Freedman L. Bayesian statistical methods (editorial). *Br Med J* 1996;313:569-70.
 2. Lilford RJ, Braunholtz. The statistical basis of public policy: a paradigm shift is overdue. *Br Med J* 1996;313:603-7.
 3. Greenland S. Probability logic and probabilistic induction. *Epidemiology* 1998;9(3):322-32.
 4. Goodman SN. Toward Evidence-Based Medical Statistics. 1: The P value fallacy. *Ann Intern Med* 1999;130:995-1004.
 5. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (I). Bayesianismo frente a frecuentismo: sus respectivas implicaciones prácticas en relación con el análisis de datos. *Med Clin (Barc)* 1996;107:467-71.
 6. Bacallao J. La perspectiva exploratorio-confirmatoria en las aplicaciones biomédicas de la estadística: dos diálogos (y II). Consideraciones críticas acerca de las pruebas de significación. *Med Clin (Barc)* 1996;107:539-43.
 7. Silva LC. Cultura estadística de investigación en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1997.
 8. Savage IR. Nonparametric statistics. *J Am Statist Assoc* 1957;52:332-3.
 9. Bakan D. The test of significance in psychological research. *Psychol Bull* 1966;66:423-37.
 10. Sackett DL. Bias in analytic research. *J Chron Dis* 1979;32:51-7.
 11. Neyman J, Pearson E. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Trans Royal Soc London* 1933;231:289-337.
 12. Rozeboom WW. The fallacy of the null hypothesis significance test. *Psychological Bulletin* 1960;56:26-47.
 13. Davidoff F. Standing Statistics Right Side Up. *Ann Intern Med* 1999;130:1019-21.
 14. Vandembrouke JP. Those who were wrong. *Am J Epidemiol* 1989;130:3-5.
 15. Efron B. Why isn't everyone a Bayesian? *American Statistician* 1986;40:1-6.
-